<u>PES UNIVERSITY</u>
<u>Data Analytics- EC campus</u>
<u>Section: F,G,I and J</u>

Format for Literature Survey Report

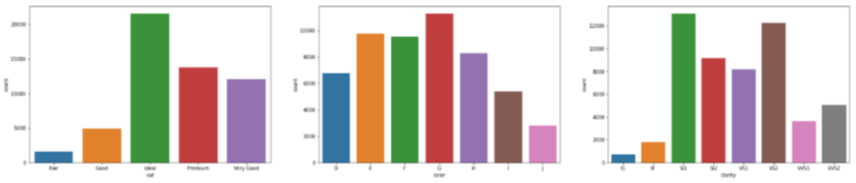
1.Project Title	Diamond Price Prediction	
2.Team Name	Regression Profession	
3.Team Members	SRN1: PES2UG20CS373	Name1: UMA KRISHNAN
	SRN2: PES2UG20CS353	Name2: Srushti V Reddy
	SRN3: PES2UG20CS365	Name3: Tanuj Sadasivam
4.Dataset used	Diamond Prices – Nancy Al Aswad on Kaggle	
5.Link for the Dataset	[Diamonds Prices | Kaggle]()	
6. Github link:	https://github.com/uma-krish/diamond-prices	

7. Problem Statement

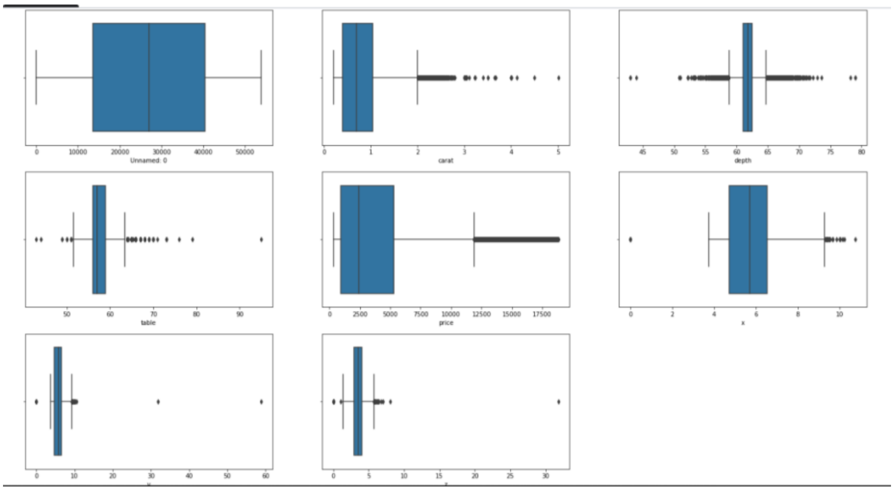
The Diamond market serves as a brilliant investment as loose diamonds are brilliant assets due to their reliability. They are reliable as their market is only modestly volatile. Thus, they are very minimally subjective to depreciation and are marginally vulnerable to inflation which generally returns good profits. Using data about diamonds and their features, we intend to analyse the price of Diamonds in the current day on the basis of Carat, Cut, Clarity, Colour, Depth, Table, their dimensions and how these features affect their price in the market.

8. EDA and Visualisation

1. Dataset contains 53943 rows and 11 columns.
2. 0 null values
3. Contains numerical and categorical datatype columns.
4. Categorical Distributions



5. Outliers - 8689



6. No inconsistencies
7. No duplicates

9. Summarise the literature survey

1. Paper 1 aims to find the most efficient and accurate model to predict the price of diamonds using a regression technique because the price is continuous. The models used to predict diamond prices are k-Nearest Neighbors (k NN) and Least Absolute Shrinkage and Selection Operator (LASSO). The process is carried out by selecting features, considering the value of k from k-NN and alpha from LASSO to ensure optimal accuracy.
2. Paper 2 aims to develop a valuation model for cut diamonds based on data published on the Internet. Regression trees (Classification and Regression Trees and Chi-Square Automatic Interaction Detection) and neural networks (using backpropagation) are used for this purpose. The proposed approaches have a complementary role in the application. Neural networks have a better performance in prediction, accounting for around 96% of cut diamond unit prices variation. The role of regression trees is fundamental in interpretability, helping to understand the contribution of predictors in pricing
3. Paper 3's main focus is to determine which gives the best accuracy between linear regression and ensemble model. After performing Chi Square, PCA and RFE we find that Random Forest Approach under ensemble learning gives the best accuracy - Gold and Diamond Price Prediction Using Enhanced Ensemble Learning.
4. Paper 4 mainly focuses on comparing the various models like Linear Regression, Random Forest Regression, Gradient Boosting Regressor, Polynomial Regression, Neural Network to get the one with the lowest error in prediction considering the noise present , we find that random forest regression gives us the best results with the lowest possible error , which is validated using cross validation - Machine Learning Algorithms for Diamond Price Prediction
5. Paper 5 aimed at, from data preprocessing, to find a correlation between the dataset attributes, train the models, test their accuracy, and analyze their outcomes, after which we determined that random forest regression model was the one with highest accuracy at 97%.
6. Paper 6 aimed at coming up with the most efficient algorithm for the price prediction of diamonds, using Non-Supervised Models. The algorithms such as K-Neighbors regression, CatBoost regression, Huber regression, Extra tree regression, Passive Aggressive regression, Bayesian Regression and XGBoost Regression were used to train the particular machine learning models on the diamond dataset for the prediction of diamond prices based on various attributes. It was concluded that CatBoost Regression model was the most efficient with an accuracy of over 98% bordering on 99%.

10. Specific Problem

Predicting the price of diamonds based on selected features

11. References

1. Cardoso, M.G.M.S. and Chambel, L. (2005), A valuation model for cut diamonds. International Transactions in Operational Research, 12: 417-436. https://doi.org/10.1111/j.1475-3995.2005.00516.x
2. S. A. Fitriani, Y. Astuti and I. R. Wulandari, "Least Absolute Shrinkage and Selection Operator (LASSO) and k-Nearest Neighbors (k-NN) Algorithm Analysis Based on Feature Selection for Diamond Price Prediction," 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022, pp. 135-139, doi: 10.1109/ISMODE53584.2022.9742936.
3. A. C. Pandey, S. Misra and M. Saxena, "Gold and Diamond Price Prediction Using Enhanced Ensemble Learning," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-4, doi: 10.1109/IC3.2019.8844910.
4. TY - BOOK AU - Alsuraihi, Waad AU - Al-hazmi, Ekram AU - Bawazeer, Kholoud AU - Alghamdi, Hanan PY - 2020/03/20 SP - 150 EP - 154 T1 - Machine Learning Algorithms for Diamond Price Prediction DO - 10.1145/3388818.3393715 ER -
5. G. Sharma, V. Tripathi, M. Mahajan and A. Kumar Srivastava, "Comparative Analysis of Supervised Models for Diamond Price Prediction," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 1019-1022, doi: 10.1109/Confluence51648.2021.9377183.

6. H. Mihir, M. I. Patel, S. Jani and R. Gajjar, "Diamond Price Prediction using Machine Learning," 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4), 2021, pp. 1-5, doi: 10.1109/C2I454156.2021.9689412.